

MetaArchive of Southern Digital Culture:

The Preservation Network

Tyler Walters
Associate Director
Digital and Technical Services

Georgia Institute of Technology Library
and Information Center

MetaArchive

Preservation Network

- Preservation Network:
 - Approach – Develop distributed preservation network infrastructure for shared preservation of digital content
 - Peer-to-peer network architecture for content preservation. Six nodes. 2.2 TB storage in aggregate. Each node communicates with all other nodes
 - All nodes serve as joint custodians of content harvested. No data lost if one node withdraws or becomes dysfunctional. Reliably preserved and validated at all preservation sites

MetaArchive

Preservation Network

- Network based technologically on LOCKSS software for distributed digital archiving and preservation, which Emory University has partnered with Stanford University and others to develop.
- The plan for this project builds on relationships and workflows developed during previous projects of the MetaScholar Initiative and other collaborating consortia

Preservation Network

Technical Features, etc.

- Decentralized Approach

(question “one copy at one inst.” approach)

- Most effective digital preservation efforts in practice succeed through some strategy for distributing copies of content in secure, distributed locations over time. This is much like the Internet itself for access

Preservation Network Technical Features, etc.

Adapting the LOCKSS Software:

- Allows Cooperative to practice its “distributed digital replication” approach. Replicate archival material, not just published material, as originally designed
 - ◆ data integrity checks
 - ◆ rigorous security checks
 - ◆ focused web crawls to gather/ingest digital content
 - ◆ problem of dynamically constructed content (to be studied)

Preservation Network

Technical Features, etc.

Migrating Digital Content:

- Digital content must be maintained in a manner that facilitates migrating it over time, across technologies
- Software used in migration needs to be open source and preserved itself
- File formats and data structures must be in accordance with standards to ensure migration
- Metadata must be standardized to migrate it as well

Preservation Network Technical Features, etc.

Dark Archiving:

- Advantage: many preservation efforts mix high accessibility online with long-term access (preservation). High accessibility equals high costs
- Content in preservation network discoverable from metadata via OAI-PMH, but downloadable only through restricted means
- Processes will be developed to retrieve items from the preservation network

Preservation Network Technical Features, etc.

Low Cost:

- Designed for minimal expenditures, low barriers to adoption, for medium-sized institutions.
- The Cooperative will develop a freely available, open source adaptation of the LOCKSS software
- Runs on inexpensive computers, modest degree of systems administration for ongoing maintenance

Preservation Network Technical Features, etc.

Metadata Registries:

- Collection registry – subset of the conspectus, i.e. collections actually harvested. Made public via OAI-PMH
- Content registry – metadata for each digital object aggregated during the harvesting process, and RSA MD5 data validation. Not public, members only
- Both registries kept on servers separate from preservation network

Preservation Network Technical Features, etc.

Flexible, adaptable multi-institutional
model:

- Purpose is to serve as a model for other consortia to utilize at low cost, simple technology to effect successful digital preservation

Cooperative Agreement Considerations

- Membership criteria – to develop a model agreement
- Roles and responsibilities – joint and equal custodians of the content harvested
- Sustainability plan
- Ensure broad applicability
- * All partners are members of the Internet 2 Consortium
- Self-sustaining incentive -- preserve institution-produced content

Types of Content

- Website exhibitions
- Digital masters (of scanned images of brittle analog originals where “rescanning” opportunities are low)
- Research databases (created by scholars working in the field, i.e. ethnographic investigations, image databases, digitally recorded interviews, etc.).
- Institutional materials – digital collections, digital intellectual output, i.e. reports, papers, learning objects, etc.

■ Questions / Comments:

Tyler Walters

Associate Director, Digital and Technical Services

404-385-4489

Tyler.Walters@library.gatech.edu

Georgia Institute of Technology Library and
Information Center

Atlanta, GA USA 30332-0900